

-95 ttaactctctggtctccgtgtctcctctcttctcctgctgcttccttttaacactctctt -36
-35 catttgccctttttgatttagatccaaagaagcagacATGTCCTCGGCGCCGTCTCCGGGG 24
M S S A P S P G
25 ACTGGTTTCGCCTCCATCTCCACCATCAAACCTCCACAACCACCACTCCTCCTCCAGCTTCC 84
T G S P P S P P S N* S T T T T P P P A S
85 GCTCCTCCTCCCACCACACCTTCTTCTCCTCCGCCGCCATCCACTATTCCGACATCTCCT 144
A P P P T T P S S P P P P S T I P T S P
145 CCTCCTTCTTCTCGCTCTACACCTTCTGCTCCTCCTCCATCTCCACCAACTCCATCTACG 204
P P S S R S T P S A P P P S P P T P S T
205 CCGGGATCTCCACCTCCTCTTCCTCAGCCGTCTCCACCCGTCCAACCTACGCCCGGATCT 264
P G S P P P L P Q P S P P A P T T P G S
265 CCACCCGCACCTGTTACTCCTCCTACTCGAAACCCTCCACCTTCAGTCCCAGGACCACCG 324
P P A P V T P P T R N P P P S V P G P P
325 TCCAATCCTTCACGCGAAGGAGGATCTCCTCGACCTCCATCTTCTCCCTCGCCGCCGTCT 384
S N P S R E G G S P R P P S S P S P P S
385 CTTTCTTCCGACGGTTTATCAACAGGAGTGGTGGTGGGAATCGCCATCGGAGGAGTCGCT 444
P S S D G L S T G V V V G I A I G G V A
445 CTGCTTGATAGTGACTCTGATTTGTCTCCTCTGTAAGAAGAAACGACGGAGAGACGAA 504
L L V I V T L I C L L C K K K R R R D E
505 GAAGATGCTTACTATGTTCCCTCCGCCACCTCCTCCTGGTCCCAAAGCCGGAGGACCTTAC 564
E D A Y Y V P P P P P P G P K A G G P Y
565 GGTGGACAGCAGCAACAATGGCGGCAACAAAACGCAACACCACCGTCAGATCATGTCGTG 624
G G Q Q Q Q W R Q Q N A T P P S D H V V
625 ACGTCACTACCACCACCACCTAAGGCTCCATCTCCACCACGGCAACCTCCTCCACCTCCA 684
T S L P P P P K A P S P P R Q P P P P P

FIG. 1C(1)

Figure 1 C

685 CCACCGCCTTTCATGAGCAGCAGCGGCGGCTCCGACTACTCGGACCGTCCAGTTCTTCCT 744
 P P P F M S S S G G S D Y S D R P V L P
 745 CCACCGTCTCCAGGGCTTGTGTTAGGCTTCTCCAAAAGCACTTTCACATACGAGGAGCTA 804
 P P S P G L V L G F S K S T F T Y E E L
 805 GCTAGAGCCACCAATGGTTTCTCCGAGGCGAACTTGTTAGGACAAGGCGGGTTCGGTTAC 864
 A R A T N ^IG F S E A N L L G Q G G F G Y
 865 GTGCACAAAGGTGTGTTGCCTAGTGGGAAAGAAGTTGCTGTGAAGCAGTTGAAAGTTGGG 924
 V H K G V L P S ^{II}G K E V A V K Q L K V G
 925 AGTGCTCAGGGAGAGAGGGAGTTTCAGGCAGAGGTTGAGATCATCAGCAGAGTTCACCAC 984
 S G ^{III}Q G E R E F Q A E V E I I S R V ^{IV}H H
 985 AGGCATCTGGTGTCTCTTGTGTTATTGCATCGCCGGTGCCAAAAGATTGCTTGTCTAT 1044
 R H L V S L V G Y C I A ^VG A K R L L V Y
 1045 GAGTTTGTTCCTAACAACAATCTCGAGCTTCACCTCCATGGCGAGGGACGGCCTACAATG 1104
 E F V P N N N L E L H L H G E G R ^{VI}P T M
 1105 GAATGGAGCACCAGATTGAAGATTGCTCTTGGATCTGCTAAAGGACTTTCTTATCTTCAT 1164
 E W S T R L K I A L G S A K G L S Y L H
 1165 GAAGATTGCAATCCTAAAATCATTACCGTGATATCAAGGCTTCAAACATATTGATAGAT 1224
 E D C N P K I I H R D I K A S N I L I D
 1225 TTCAAGTTTGAAGCTAAGGTTGCTGATTTTGGTCTTGCTAAGATTGCTTCTGATACAAAC 1284
 F K F ^{VII}E A K V A D F G L A K I A S D T N
 1285 ACGCATGTATCAACACGTGTGATGGGAACCTTTGGGTACTTGGCTCCGGAATACGCTGCA 1344
 T H V ^{VIII}S T R V M G T F G Y L A P E Y ^{IX}A A
 1345 AGCGGAAAGCTCACGGAGAAGTCTGACGTTTTCTCATTTGGCGTTGTGCTTTTGGAGCTC 1404
 S G K L T E K S D V F S F G V V L L E L

FIG. 1C(2)

1405 ATTACTGGACGTCGACCCGTTGATGCCAACAATGTCTATGTAGATGACAGCTTAGTTGAC 1464
I T G R R P V D A N N^X V Y V D D S L V D
 1465 TGGGCACGACCATTGCTTAACCGAGCATCTGAGCAAGGAGACTTTGAGGGTTTAGCTGAT 1524
W A R P L L N R A S E Q G D F E G L A D
 1525 GCAAAGATGAATAATGGGTATGACAGAGAGGAGATGGCTCGCATGGTTGCTTGTGCTGCG 1584
A K M N N G Y D R E^{XI} E M A R M V A C A A
 1585 GCTTGTGTTTCGCCATTCAGCTCGCCGAGACCTCGCATGAGCCAGATTGTGCGTGCGTTA 1644
A C V R M S A R R R P R M S Q I V R A L
 1645 GAAGGAAATGTATCACTGTCAGATCTTAACGAAGGGATGAGACCAGGTCAAAGCAATGTA 1704
 E G N* V S L S D L N E G M R P G Q S N V
 1705 TACAGCTCATACGGAGGAAGCACCGATTATGACTCGAGCCAGTACAATGAAGACATGAAG 1764
 Y S S Y G G S T D Y D S S Q Y N E D M K
 1765 AAGTTTAGGAAAATGGCACTTGGAAGTCAAGAGTACAACGCCACGGGTGAGTACAGTAAT 1824
 K F R K M A L G T Q E Y N* A T G E Y S N
 1825 CCGACCAGTGACTATGGACTGTACCCGTCTGGTTCAAGCAGCGAGGGCCAAACCACACGC 1884
 P T S D Y G L Y P S G S S S E G Q T T R
 1885 GAAATGGAGATGGGGAAGATTAAGAGAACCGGTCAGGGTTATAGTGGACCTTCTCTTTAA 1944
 E M E M G K I K R T G Q G Y S G P S L *
 1945 accagatgggagagaaaattgaaggggtgttttttcattattttttaaaactgtaaagata 2004
 2005 tgagaaaattgccttactctaattaaaaccactacgatataagggttataatacgtttttga 2064
 2065 attggttttttaaaaaaaaaaaaaaaaaaaaaa 2093

FIG. 1C(3)

10	20	30	40	50	60
TTAACTCTCT	GGTCTCCGTG	TCTCCTCTCT	TCTCCTGCTG	CTTCCTTTTA	ACACTCTCTT
70	80	90	100	110	120
CATTTGCCTT	TTTGATTAG	ATCCAAAGAA	GCAGACATGT	CCTCGGCGCC	GTCTCCGGGG
130	140	150	160	170	180
ACTGGTTCGC	CTCCATCTCC	ACCATCAAAC	TCCACAACCA	CCACTCCTCC	TCCAGCTTCC
190	200	210	220	230	240
GCTCCTCCTC	CCACCACACC	TTCTTCTCCT	CCGCCGCCAT	CCACTATTCC	GACATCTCCT
250	260	270	280	290	300
CCTCCTTCTT	CTCGCTCTAC	ACCTTCTGCT	CCTCCTCCAT	CTCCACCAAC	TCCATCTACG
310	320	330	340	350	360
CCGGGATCTC	CACCTCCTCT	TCCTCAGCCG	TCTCCACCCG	CTCCAACCTAC	GCCCGGATCT
370	380	390	400	410	420
CCACCCGCAC	CTGTTACTCC	TCCTACTCGA	AACCCTCCAC	CTTCAGTCCC	AGGACCACCG
430	440	450	460	470	480
TCCAATCCTT	CACGCGAAGG	AGGATCTCCT	CGACCTCCAT	CTTCTCCCTC	GCCGCCGTCT
490	500	510	520	530	540
CCTTCTTCCG	ACGGTTTATC	AACAGGAGTG	GTGGTGGGAA	TCGCCATCGG	AGGAGTCGCT
550	560	570	580	590	600
CTGCTTGGA	TAGTGACTCT	GATTTGTCTC	CTCTGTAAGA	AGAAACGACG	GAGAGACGAA
610	620	630	640	650	660
GAAGATGCTT	ACTATGTTCC	TCCGCCACCT	CCTCCTGGTC	CCAAAGCCGG	AGGACCTTAC
670	680	690	700	710	720
GGTGGACAGC	AGCAACAATG	GCGGCAACAA	AACGCAACAC	CACCGTCAGA	TCATGTCGTG
730	740	750	760	770	780
ACGTCACTAC	CACCACCACC	TAAGGCTCCA	TCTCCACCAC	GGCAACCTCC	TCCACCTCCA
790	800	810	820	830	840
CCACCGCCTT	TCATGAGCAG	CAGCGGCGGC	TCCGACTACT	CGGACCGTCC	AGTTCTTCCT
850	860	870	880	890	900
CCACCGTCTC	CAGGGCTTGT	GTTAGGCTTC	TCCAAAAGCA	CTTTCACATA	CGAGGAGCTA
910	920	930	940	950	960
GCTAGAGCCA	CCAATGGTTT	CTCCGAGGCG	AACTTGTTAG	GACAAGGCGG	GTTTCGGTTAC
970	980	990	1000	1010	1020
GTGCACAAAG	GTGTGTTGCC	TAGTGGGAAA	GAAGTTGCTG	TGAAGCAGTT	GAAAGTTGGG
1030	1040	1050	1060	1070	1080
AGTGGTCAGG	GAGAGAGGGA	GTTTCAGGCA	GAGGTTGAGA	TCATCAGCAG	AGTTCACCAC
1090	1100	1110	1120	1130	1140
AGGCATCTGG	TGTCTCTTGT	TGGTTATTGC	ATCGCCGGTG	CCAAAAGATT	GCTTGTCTAT
1150	1160	1170	1180	1190	1200
GAGTTTGTTT	CTAACAACAA	TCTCGAGCTT	CACCTCCATG	GCGAGGGACG	GCCTACAATG
1210	1220	1230	1240	1250	1260
GAATGGAGCA	CCAGATTGAA	GATTGCTCTT	GGATCTGCTA	AAGGACTTTC	TTATCTTCAT
1270	1280	1290	1300	1310	1320
GAAGATTGCA	ATCCTAAAAT	CATTACCGGT	GATATCAAGG	CTTCAAACAT	ATTGATAGAT
1330	1340	1350	1360	1370	1380
TTCAAGTTTG	AAGCTAAGGT	TGCTGATTTT	GGTCTTGCTA	AGATTGCTTC	TGATACAAAC
1390	1400	1410	1420	1430	1440
ACGCATGTAT	CAACACGTGT	GATGGGAACC	TTTGGGTACT	TGGCTCCGGA	ATACGCTGCA
1450	1460	1470	1480	1490	1500
AGCGGAAAGC	TCACGGAGAA	GTCTGACGTT	TTCTCATTTG	GCGTTGTGCT	TTTGGAGCTC
1510	1520	1530	1540	1550	1560

FIG. 1D(1)

```

ATTACTGGAC GTCGACCCGT TGATGCCAAC AATGTCTATG TAGATGACAG CTTAGTTGAC
1570 1580 1590 1600 1610 1620
TGGGCACGAC CATTGCTTAA CCGAGCATCT GAGCAAGGAG ACTTTGAGGG TTTAGCTGAT
1630 1640 1650 1660 1670 1680
GCAAAGATGA ATAATGGGTA TGACAGAGAG GAGATGGCTC GCATGGTTGC TTGTGCTGCC
1690 1700 1710 1720 1730 1740
GCTTGTGTTT GCCATTGAGC TCGCCGCAGA CCTCGCATGA GCCAGATTGT GCGTGCCTTA
1750 1760 1770 1780 1790 1800
GAAGGAAATG TATCACTGTC AGATCTTAAC GAAGGGATGA GACCAGGTCA AAGCAATGTA
1810 1820 1830 1840 1850 1860
TACAGCTCAT ACGGAGGAAG CACCGATTAT GACTCGAGCC AGTACAATGA AGACATGAAG
1870 1880 1890 1900 1910 1920
AAGTTTAGGA AAATGGCACT TGGAACCTAA GAGTACAACG CCACGGGTGA GTACAGTAAT
1930 1940 1950 1960 1970 1980
CCGACCATG ACTATGGAT GTACCCGTCT GGTTCAGCA GCGAGGGCCA AACCACACGC
1990 2000 2010 2020 2030 2040
GAAATGGAGA TGGGGAGAT TAAGAGAACC GGTGAGGTT ATAGTGGACC TTCTCTTTAA
2050 2060 2070 2080 2090 2100
ACCAGATGGG AGAGAAATTG AAGGGTGTTC TTTCATTATT TTTTAAAC TGTAAGATA
2110 2120 2130 2140 2150 2160
TGAGAAATT GCCTTACTCT AATTAAACC ACTACGATAT AAGGTTATAA TACGTTTGA
2170 2180 2190 2200 2210 2220
ATTGGTTTTT AAAAAAAAAA AAAAAAAAAA .....

```

Figure 1(d) (continued)

```

LTLWSPCLLS SPAASF*HSL HLPF*FRSKE ADMSSAPSPG TGSPPSPPSN STTTTPPPAS
APPTTPSSP PPPSTIPTSP PPSSRSTPSA PPSPTTPST PGSPPPLPQP SPPAFTTPGS
PPAPVTPPTR NPPPSVPGFP SNPSREGGSP RPPSSPSPPS PSSDGLSTGV VVGIAIGGVA
LLVIVTLICL LCKKKRRRDE EDAYYVPPFP PPGPKAGGPY GGQQQWRQO NATPPSDHVV
TSLPPPPKAP SPPRQPPPPP PPFMSSSGG SDYSDRPVLP PPSPLVLGF SKSTFTYEEL
ARATNGFSEA NLLGQGGFGY VHKGVLPSGK EVAVKQLKVG SGQGEREFOA EVEIISRVBH
RHLVSLVGYC IAGAKRLLY EFVFNNNLEL HLHGEGRPTM EWSTRLKIAL GSAKGLSYLH
EDCNPKIIHR DIKASNILID FKFEAKVADF GLAKIASDTN THVSTRVMGT FGYLEPEYAA
SGKLTEKSDV FSFGVVLLEL ITGRRPVDAN NVYVDDSLVD WARPLLNRAS EQGDFEGLAD
AKMNNGYDRE EMARMVACAA ACVRHSARRR PRMSQIVRAL EGNVSLSDLN EGMRFQGSNV
YSSYGGSTDY DSSQYNEDMK KERKMALGTQ EYNATGEYSN PTSDYGLYPS GSSSEGQTTT
EMEMGKIKRT GQGYSGPSL TRWERN*RVF FHYFFKTVKI *ENCLTIKT TTI*GYNTF*
IGF*KKKKK

```

Figure 1(e)

ATTACTGGAC	GTCGACCCGT	TGATGCCAAC	AATGTCTATG	TAGATGACAG	CTTAGTTGAC
1570	1580	1590	1600	1610	1620
TGGGCACGAC	CATTGCTTAA	CCGAGCATCT	GAGCAAGGAG	ACTTTGAGGG	TTTAGCTGAT
1630	1640	1650	1660	1670	1680
GCAAAGATGA	ATAATGGGTA	TGACAGAGAG	GAGATGGCTC	GCATGGTTGC	TTGTGCTGCG
1690	1700	1710	1720	1730	1740
GCTTGTGTTT	GCCATTTCAGC	TCGCCGCAGA	CCTCGCATGA	GCCAGATTGT	GCGTGCGTTA
1750	1760	1770	1780	1790	1800
GAAGGAAATG	TATCACTGTC	AGATCTTAAC	GAAGGGATGA	GACCAGGTCA	AAGCAATGTA
1810	1820	1830	1840	1850	1860
TACAGCTCAT	ACGGAGGAAG	CACCGATTAT	GACTCGAGCC	AGTACAATGA	AGACATGAAG
1870	1880	1890	1900	1910	1920
AAGTTTAGGA	AAATGGCACT	TGGAAC TCA	GAGTACAACG	CCACGGGTGA	GTACAGTAAT
1930	1940	1950	1960	1970	1980
CCGACCAGTG	ACTATGGACT	GTACCCGTCT	GGTTCAAGCA	GCGAGGGCCA	AACCACACGC
1990	2000	2010	2020	2030	2040
GAAATGGAGA	TGGGGAAGAT	TAAGAGAACC	GGTCAGGGTT	ATAGTGGACC	TTCTCTTTAA
2050	2060	2070	2080	2090	2100
ACCAGATGGG	AGAGAAATTG	AAGGGTGTTT	TTTCATTATT	TTTTTAAAC	TGTAAAGATA
2110	2120	2130	2140	2150	2160
TGAGAAAATT	GCCTTACTCT	AATTAACACC	ACTACGATAT	AAGGTTATAA	TACGTTTTGA
2170	2180	2190	2200	2210	2220
ATTGTTTTTT	AAAAAAAAAA	AAAAAAAAAA

FIG. 1D(2)

LTWSPCLLS	SPAASF*HSL	HLPF*FRSKE	ADMSSAPSPG	TGSPPSPPSN	STTTTPPPAS
APPPTTPSSP	PPPSTIPTSP	PPSSRSTPSA	PPPSPTPST	PGSPPPLPQP	SPPAPTTPGS
PPAPVTPPTR	NPPPSVPGPP	SNPSREGGSP	RPPSSPSPPS	PSSDGLSTGV	VVGIAIGGVA
LLVIVTLICL	LCKKKRRRDE	EDAYYVPPPP	PPGPKAGGPY	GGQQQQWRQQ	NATPPSDHVV
TSLPPPPKAP	SPPRQPPPPP	PPPFMSSSGG	SDYSDRPVLP	PPSPGLVLGF	SKSTFTYEEL
ARATNGFSEA	NLLGQGGFGY	VHKGVLPSGK	EVAVKQLKVG	SGQGEREFQA	EVEIISRHH
RHLVSLVGYC	IAGAKRLLVY	EFVPNNNLEL	HLHGEGRPTM	EWSTRLKIAL	GSAGLSYLH
EDCNPKIIHR	DIKASNILID	FKFEAKVADF	GLAKIASDTN	THVSTRVMGT	FGYLAPEYAA
SGKLTEKSDV	FSFGVVLEL	ITGRRPVDAN	NVYVDDSLVD	WARPLLNRAS	EQGDFEGLAD
AKMNNGYDRE	EMARMVACAA	ACVRHSARRR	PRMSQIVRAL	EGNVSLSDLN	EGMRPGQSNV
YSSYGGSTDY	DSSQYNEDMK	KFRKMALGTQ	EYNATGEYSN	PTSDYGLYPS	GSSSEGQTR
EMEMGKIKRT	GQGYSGPSL	TRWERN*RVF	FHYFFKTVKI	*ENCLT LIKT	TTI*GYNTF*
IGF*KKKKK					

FIG. 1E